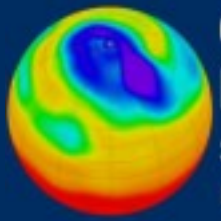
Science & Technology Facilities Council

Vocabularies for climate data

Martin Juckes

- Models, observations and reanalysis: an integrated approach;
- Vocabularies for CMIP



Centre for Environmental Data Archival

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL

NATURAL ENVIRONMENT RESEARCH COUNCIL

Climate Information Platform for Copernicus

Helping Europe respond to the impact of climate change

A new Climate data infrastructure for Europe – first steps

CLIPC is a 3 year research project, developing a prototype portal for the Copernicus Climate Change Service

www.clipc.eu
Abingdon, February, 2015

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

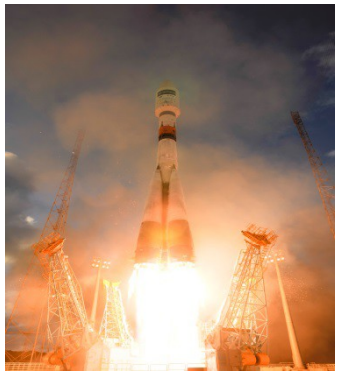
CLIPC and the Copernicus programme

FP7 and H2020

Copernicus

Users

CLIPC is a research project – not a component of the Copernicus programme



The test for CLIPC components will be user evaluation

IT components

Standards

Enhanced data

Future research

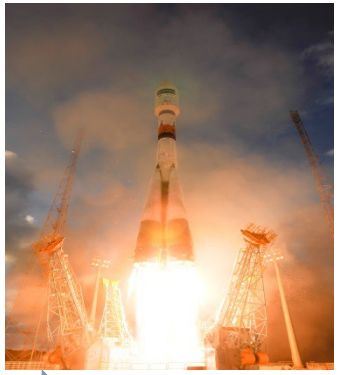
People, networks

Copernicus Climate Change Service (C3S)

European Environment Agency

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies and Catalogues

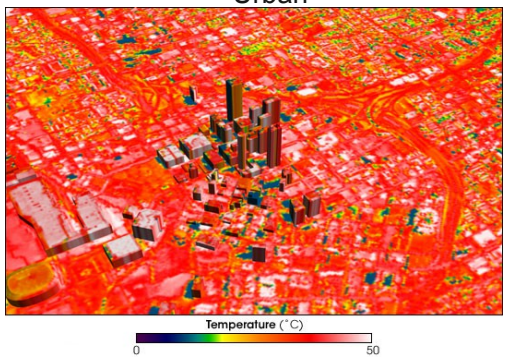


Integration across science domains

Integration through the processing chain

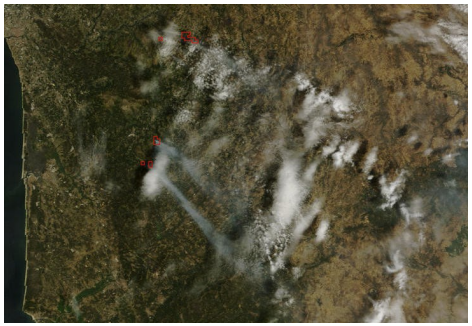
Integration across sectors

Urban



Thermal image of Atlanta (US) (Wikipedia)

Rural



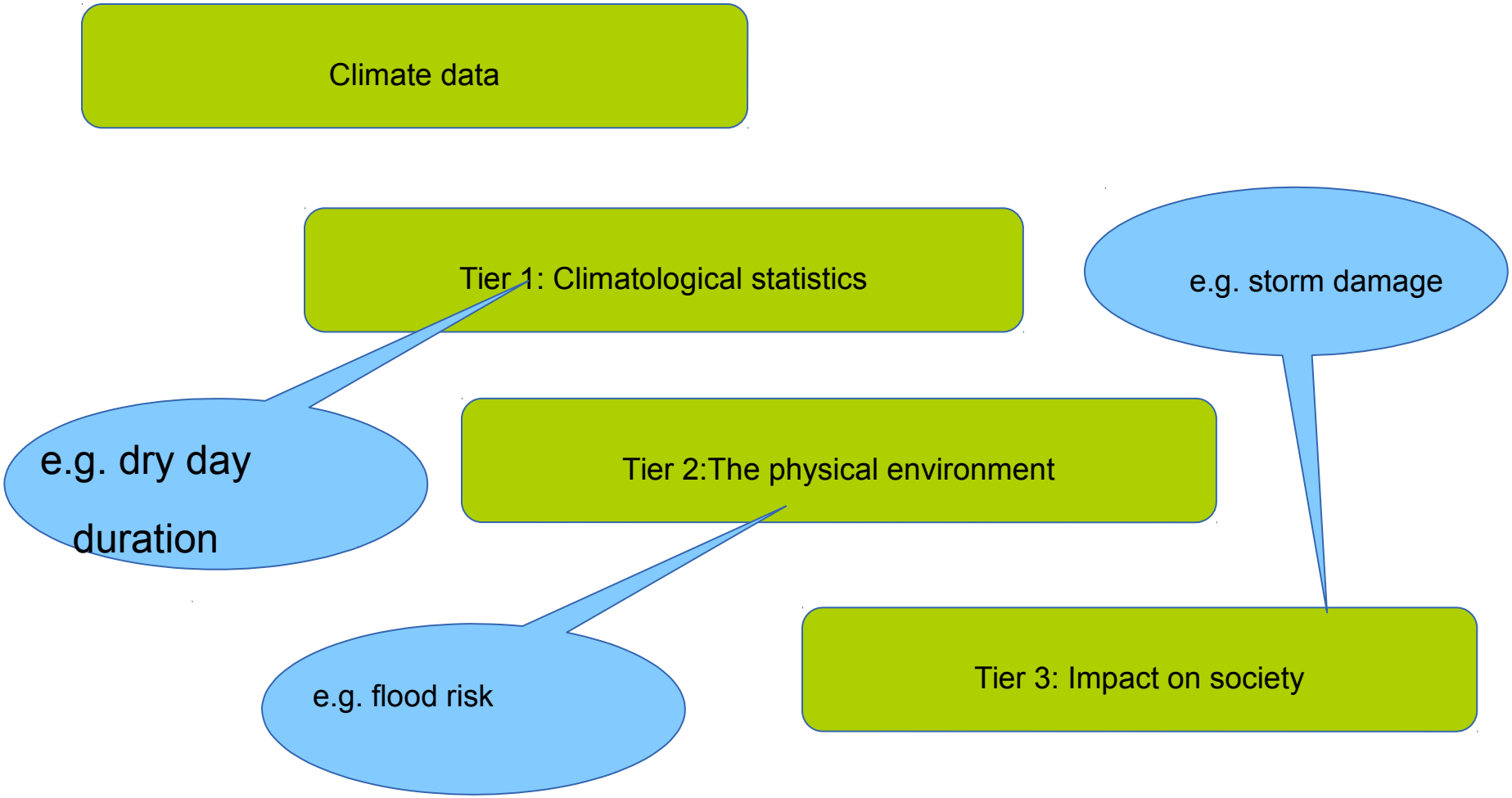
2013 fires (Portugal) (Uni. Freiburg)

Water

Elbe river (Germany) (Wikipedia)

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Climate data and 3 “tiers” of indicators



This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies: objectives

- Clear terminology to support
  - management,
  - curation,
  - discovery and
  - use of data;
- Traceable references for data: ease of use and citation metrics;

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies: contrasting examples

European Space Agency (ESA) Climate Change Initiative (CCI)

- Programme (ESA-CCI)
- CCI Project (ECV)
- Data Type (Variable name)
- Product (variations in scientific and technical process)
- Additional Seg. (extendable, structured term)
- Indicative time
- Data specification version, file version

World Climate Research Programme (WCRP) Coupled Model Intercomparison Project (CMIP)

- Activity/Institute/Model/Experiment/Ensemble
- Realm/Variable name
- Frequency/Table
- Start and end dates

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

	CCI	CMIP5
Strengths	• Clear differentiation between specification of physical entity and data structure; • Versioning of data specification supported;	• Terms are explicitly defined or have simple numerical structure: very easy to interpret in software.
Weaknesses	• Complex, science team specific terms in file names; • Algorithm used in creating data is of high scientific importance but not clearly labelled;	• Vocabularies mix specification of physical entities with specification of shape of data; • No clear versioning of specifications; • Ambiguity in physics and initialisation methods.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies: categories of categories

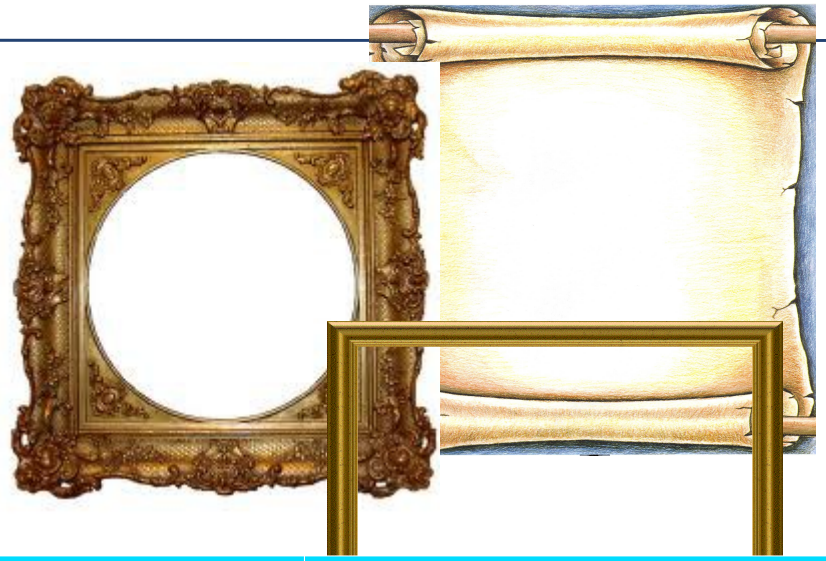
By governance

All vocabularies are equal, but some are more equal than others.

Governance	Description	Examples
Standards	Clear and effective governance; terms added to meet community requirements.	ECVs, CF standard names.
Registries	Terms added to suit needs of specialist teams: registered to avoid duplicates and make definitions available. Sometimes with guidance on styling of names.	ESA CCI Product.
Extensible	Allow individual data providers to add terms within a specified pattern.	Ensemble number; file version

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies and Catalogues



Objective	Description	Examples
View (many worlds)	There are many estimates of the state of the world, and, from simulations, many realisations of possible worlds.	Model, experiment, ensemble; mission, analysis method;
Focus (place or property within the world)	In each scientific domain there are a variety of ways of identifying focussed areas of information within an overall view of the world;	Realm, variable, ECV.
Style (presentation)	Data can be presented in different ways, varying sampling rate, file format,etc.	Frequency; ESA processing level.

Managing vocabularies and their relationships

E.g. GRIB and CF are externally maintained: CLIPC will maintain a set of relationships

Maintenance:review and approve relationships, record provenance of terms and relationships.

Core vocabulary services: BODC vocab server (SKOS)

User interface: search exploiting semantic relationships

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

GRIB & NetCDF

Foundation terms *Stable terms defined through transparent governance mechanism.*	**Common terms** *Corresponding to quantities the scientists need – defined through a collection of attributes*.
CF standard names	MIP variables
GRIB/WMO code tables	GRIB API
Term A =/>/< Term B	{Term A, att1=X, att2=Y, ..} = {Term B, att3=Z, att4=W,...}

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Categories of terms

CMIP (NetCDF) Realms	GRIB Domains	GCOS Categories	Aristotle Elements
•Atmosphere •Ocean •Land •Land Ice •Sea Ice •aerosol •Atmospheric Chemistry •Ocean Biogeochemistry	•Meteorology •Oceanography •Hydrology •Land Surface •Space Weather •Other Space	•Atmospheric ➔Surface ➔Upper air ➔Composition •Terrestrial •Oceanic ➔Surface ➔Subsurface	• Air • Earth • Water • Fire

CMIP “atmosphere” ≠ GRIB “Meteorology” ≠ GCOS “Atmospheric” ≠ Aristotle “Air”

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Summary and outlook

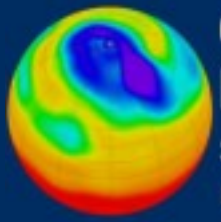
- Many different vocabularies are in use to organise climate information;
- By organising the vocabularies in governance and objective classes we can create a structured and flexible reference system;
- Climate vocabularies tend to be internalised: not exported to truly interdisciplinary systems (e.g. DOI dictionary of terms);
- A more generic approach to reference vocabularies will support integration into the knowledge base.

	Standard	Registry	Extensible
View	CMIP experiment;	Model; algorithm	Ensemble; Version
Focus	Essential Climate Variables; modelling realm;	MIP variable;	
Style	CMIP frequency;	**Not needed?**	

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607418

Vocabularies for CMIP6

- Defining new variables
- Models, experiments and forcings



Vocabularies for CMIP6

- Consistently named variables for simulations and analysis by ~24 intercomparison projects;
- New variables for dynamic ice-shelves (polar stereographic);
- Data on multiple land-use tiles;
- Offline models integrated into analysis;
- Monthly averaging with time-varying masks;
- New standard names;

Models, experiments and forcings

- Many models come with family connections – naming of various family members was not consistent across CMIP5 models;
- With an increasing range of experiments (around 200), improved clarity around experiment names is needed;
- Huge array of forcing variations (sensitivity experiments forced with artificially perturbed SST and various combinations of greenhouse gasses).

Need linked information about vocabulary terms – cannot rely on “self-explanatory” terminology.